The Impasse between the Design and Evolution of Life

Philip R. Page*

Abstract

The design explanation of the origin and structure of life, which is a rational probabilistic mode of explanation, is currently sustained in tandem with the chance explanation, usually based on Darwinian polyphyletic evolution. It is hence possible for two logically incompatible explanations to be adopted concurrently by the scientific community. Evolution partially developed in response to the need for a causal story.

Keywords: design, chance, evolution, life

1 Prolegomenon

Consider a science fiction fable. In the twenty second century human spaceships reached the planet Germanicon where a fully operational calculator based on germanium technology was discovered. It could perform many of the mathematical and statistical operations known on twenty second century earth, and also numerous as yet unknown operations. Because the calculator was apparently performing mathematical and statistical operations intelligible to man, scientific investigation of its origin and structure was possible and began. After a century of investigation, the structure of the calculator was found to be fully determined by the settings of certain sites, or "nodes". There was a *thousand* nodes, each of which

* *E-mail: prp@deuteron.lanl.gov. Work location: Nuclear Physics, Theoretical Division, Los Alamos National Laboratory, Los Alamos, USA. Address: 5040 Carriage House Dr., Los Alamos, NM 87544, USA.*

could be set to *ten* possible values. The values were randomly changed a million trillion times and none of the corresponding calculators were found to be operational. The values were then reset to what they were originally. Because of the extraordinary low probability of the calculator being operational when assembled by chance, the scientists inferred that it was designed. The design inference enabled further scientific investigation along the lines of reverse engineering to fully understand the details of how the calculator was designed, yielding fecund insights along the way, and spurring new technologies back on earth.

However, as the centuries progressed the need for a causal story became pressing: Who was the designer? In the thirty second century, a thousand years after the discovery of the calculators, space exploration reached distances vastly larger that the distance between Germanicon and earth, and no plausible designer was found. Seventy planets suitable for germanium–based calculators were discovered. The planets were suitable in the sense that the nodes can be expected to assemble only once early in the history of the planet, and that if the nodes assemble in a sequence corresponding to an operational calculator, the calculator would be operational.

Although reverse engineering based on the design explanation was still a fruitful scientific enterprise, the need for a causal story, and the apparent lack of a designer, engendered the idea that the calculator could have arisen by chance. This could be achieved if scientific knowledge about the physical world is sufficient to explain the origin and structure of the calculator by purely naturalistic means. In the same century an additional, derelict, calculator was discovered, that although no longer actively operating, was in principle capable of operation. The derelict calculator contained only *five hundred* nodes. A naturalistic theory provided a plausible account of how the operational calculator could have evolved from the derelict one. Because the evolutionary naturalistic theory used the derelict calculator as a starting point, which has a significantly lower probability of arising by chance, due to the fewer nodes, this theory could conceivably challenge the design explanation. The increasing knowledge gleaned from the reverse engineering enterprise was reinterpreted in the framework of this theory, in an attempt to provide a causal story.

However, partially due to incomplete information due to the derelict calculator being in a state of decay, the evolutionary theory was not demonstrated definitively, at least up to the thirty fifth century. During those three centuries design and chance hence coexisted as modes of explanation for the intriguing operational calculator of Germanicon.

2 The design inference

As the reader might have noticed, the preceding science fiction fable is allegorical, representing the progress of the biology of the origin of life throughout history. Simply associate "Germanicon" with "earth", "calculator" with "organism", "node" with "DNA base pair" (or more generally "biological component"), "operational" with "functional" (or "living") and "derelict" with "fossilized".

On discovering that the operational calculator had a thousand nodes, each which could register ten values, a calculation shows that there are 10^{1000} possible different calculators, a fantastically large number. Given that the values were randomly changed a million trillion times, and no operational calculators were found, one is led to conclude that the operational calculator of Germanicon could not have arisen by chance and hence infer that it was designed. The design inference has recently been formalized by William Dembski [1], and we shall highlight aspects of his work specifically relevant to our considerations.

The design inference rests on the law of small probabilities: *Events with small probability do not occur by chance*, as justified by Dembski. If we identify an event as "finding an operational calculator", then the fact that for each operational calculator a million trillion non–operational calculators were found, means that the probability of the event[1] p_o is less than $\frac{1}{2}/10^{18} = \frac{1}{2}\times 10^{-18}$. (So that in a million trillion tries, the probability is less than 50% to find an operational calculator; consistent with what was observed). Since the probability of the event is small, one infers from the law of small probabilities that the event did not happen by chance. Another logical possibility is that the event occurred due to a regularity, so that the event will (almost) always happen [1], e.g. due to a necessary process or a deterministic law (or model) of nature. Dembski defines design as the logical complement of chance and regularity, so that if an event did not happen by chance or regularity, it was designed[2]. We shall not be concerned with the possibility of explaining events due to regularity, since this mode of explanation is not mentioned in the fable, and will not be important for the discussion of the origin of life. However, we shall contrast regularity and

[1]From the fable we know that there were 70 planets found which are suitable for germanium–based calculators. The correct calculation therefore gives that $p_o < \frac{1}{2}\times 70\times 10^{-18} < 10^{-16}$.

[2]This is a restricted use of the word "design" which we shall adopt throughout. In this restricted sense, a regular or chance event is not designed. However, a more general use of the term may suggest to some that e.g. the laws of nature were designed.

chance in sections 6–7. Design is henceforth simply referred to as the logical complement of chance. We hence conclude that the operational calculator of Germanicon was designed. Because design is the logical complement of chance, it is logically inconsistent to hold both modes of explanation.

It is incumbent on us to point out a common misconception: Because there are 10^{1000} possible calculators, the probability of this specific operational calculator to arise by chance is 10^{-1000}. One may easily assume that this is the appropriate probability to infer design from. However, one should not marvel at the fact that a specific calculator has been discovered, but that an operational calculator has been discovered. It is not the discovery that is exciting, but the attribute that something meaningful has been discovered, i.e. that the calculator performs known mathematical and statistical functions, and therefore is operational as a calculator. This attribute of the calculator is an example of what Dembski calls a "specified" event, and the law of small probabilities should only be applied to such events. Amongst the 10^{1000} possible calculators, there may be more operational calculators than the specific one found. Instead of the vanishingly small probability 10^{-1000}, the relevant probability to infer design from is the probability of finding an operational calculator, p_o.

A central feature of a specified event, among others, is that there is some independent side information connected to the event [1]. The knowledge of different possible mathematical and statistical functions is side information that is quite independent from the calculator of Germanicon. The fact that the calculator actually performs these functions is what is meaningful. If there is no side information connected to an event, the event is not meaningful, and there is no reason to ascribe design to the event. It is for this reason that the design argument can only be applied to specified events. In this work "specified" means "operational" in the fable. Note that there is no need for the side information to be complex (knowledge of only one mathematical or statistical function would already be useful). However, the complexity (the number of nodes and their values) of the calculator has a bearing on how p_o is calculated, determining whether design or chance is inferred.

Now let us apply the same reasoning to the origin of life. For biological organisms, the side information can be as simple as knowledge of various different possible functions, e.g. reproductive capability. By "specified" we shall mean "functional" in biology. Side information does not need to be available about complicated details concerning the functions of various organs inside the organism. The various functions of the organism should enable the organism to survive, i.e. to live.

Consider a fairly complex organism with a *billion* DNA base pairs, each of which can take *four* possible values: adenine, cytosine, guanine and thymine. Analogous to the calculators, the probability for an organism to have a specific sequence of base pairs is $4^{-1000000000} \approx 10^{-600000000}$. As before, this vanishingly small probability cannot be the basis for inferring design. The relevant probability is the probability for an organism to be functional given an *ad hoc* sequence of base pairs p_f. Although it is not our objective to estimate p_f from current scientific knowledge, we highlight some of the issues which enter into such an estimate.

Our intuition is that p_f must be tiny based on our experience that if the components of a man–made machine are randomly interchanged, one expects the machine to be inoperative. Experiments where a few base pairs are randomly changed in a living organism (called "mutation") indicate that in most cases the resulting organisms cannot be regarded as living, in the sense that they would not be able to survive, e.g. under natural selection due to the physical environment. On the other hand, random mutation due to cosmic rays has happened throughout history, and often yield (almost) imperceptible changes to an organism. If the results are so ambiguous when only a few base pairs are changed, the machine analogy leads one to expect worse results when many base pairs are randomly changed.

Also, there is nothing special about the fact that life originated on earth. If it originated on another suitable planet and intelligent life arose asking what the probability of functional organisms is, they would be in the same position as earthlings. This is called the *anthropic principle*. Since the number of possible circumstances in which life originated equals the number of suitable planets N_{pl} on which intelligent life arose[3], the probability of finding functional organisms

$$p_f = N_{pl}\, p_l,$$

where p_l is the probability of an organism to be a functional (living) organism *on earth*, given an *ad hoc* sequence of base pairs. N_{pl} is bounded above by the number of planets in the known universe, estimated to be 10^{20}.

If the probability for an organism to be functional, $N_{pl}\, p_l$, is less than $\frac{1}{2}$ one infers

[3]It is technically more correct not only to calculate the probability of life arising on *one* planet, but also on two, three, etc. However, a calculation shows that because p_l is tiny, the latter possibilities are negligible.

that the organism was designed. (The magical number "$\frac{1}{2}$" will be discussed in section 7.) If the probability for an organism to be functional is larger than $\frac{1}{2}$, one infers that the organism arose by chance. In the next section some further moves to strengthen the chance explanation will be indicated.

In this section we provided a *simpliciter* (simplified) understanding of the biology involved, which we shall refer to as such. This means that the one billion base pairs assemble randomly only once early in the history of a suitable planet.

The primary goal is an explication of the issues involved, not an exhaustive characterization: The design argument presented is a sub-case of the full argument. Other issues to address include how the constructs common to life, e.g. nucleotides, amino acids, RNA and DNA, arose in the first place and how some organismic functions can be so complex (e.g. the irreducibly complex biological machines Behe found in living organisms). Although the feedback mechanisms of biological organisms further complicate their analysis, it is still useful to consider different possible organisms by only considering the different possible base pairs, as we have done. The analogy in the fable is that although the calculator is bound to have many wires connecting the nodes, the calculator was assumed to be fully determined by the settings of the nodes. Instead of arguing for design by virtue of functionality (surviving living organisms performing various functions), one can even consider architecture (organisms possess engineering and aesthetical architectures).

3 The chance explanation

In the science fiction fable a plausible causal story is found that explains how the operational calculator with a thousand nodes could have evolved from a derelict calculator that is in principle capable of operation with five hundred nodes. What remains to be explained is how the derelict calculator arose. To understand what is at stake let us assume for the moment the same fable, but substituting a derelict calculator consisting of only two nodes. As the derelict calculator is not operational, it is not possible to randomly change the values of the nodes in order to estimate the relevant probability p_o that an arbitrary calculator is operational. However, there is something non–trivial that one can deduce. Since there are $10^2 = 100$ possible calculators and one of them is already known to be operational (in principle), and there are 70 suitable planets, $p_o \geq \frac{70}{100}$. Because it is not true that

$p_o < \frac{1}{2}$ we cannot infer design and the chance explanation prevails. We hence use chance to explain how the derelict calculator arose and then the plausible causal story of how the operational calculator arose by evolving from the derelict calculator. Assuming that "plausible" corresponds to a probability near unity, chance fully explains the operational calculator of Germanicon.

In our actual fable, the number of nodes of the derelict calculator is five hundred instead of two, so that chance does not unambiguously explain the operational calculator (as an analogous calculation shows). The gist is, however, that as we moved from a thousand to five hundred nodes, with a plausible causal story connecting them, progress towards the chance explanation was made.

Translating to biology, the causal story is the gradual or punctuated polyphyletic (between phyla) Darwinian evolution from a fossilized progenitor organism to a modern functional organism. In Appendix A we give a mathematical example to explain how evolution happens. Let's call the probability (plausibility) of the causal story p_{ev}. More precisely, p_{ev} is the probability of a potentially functional progenitor organism to evolve into a functional modern organism. Hence, p_{ev} quantifies the efficacy of evolution. (Note that p_{ev} is *not* the probability of a specific progenitor to evolve to a specific modern organism.) Assuming that evolutionary theory can provide a causal story connecting the smallest fossilized progenitor, with say a *million* base pairs of DNA, with the modern organism with a billion base pairs considered previously, the probability for the progenitor to evolve to the modern organism is

$$p_e \geq N_{pl}\, p_{ev}\, 4^{-1000000},$$

where $4^{-1000000} = 10^{-600000}$.

To summarize, the *evolutionary* understanding is as follows. A million base pairs assemble randomly only once early in the history of a suitable planet. If the corresponding organism is functional, this progenitor organism then has a chance to evolve via gradual or punctuated polyphyletic Darwinian evolution to a modern functional organism with a billion base pairs.

4 Probabilistic resources and the chance hypothesis

In section 2 our *simpliciter* understanding indicated that the probability of an organism to be functional given an *ad hoc* sequence of base pairs, the relevant probability needed in order to successfully infer the design of life, is $p_f = N_{pl}\, p_l$. Here N_{pl} is the number of suitable planets on which intelligent life arose, bounded above by 10^{20}. p_l is the probability of an organism to be functional *on earth* given an *ad hoc* sequence of base pairs. A successful design inference requires that the relevant probability be less than $\frac{1}{2}$. As current astrobiological and biological knowledge do not allow determination of either N_{pl} or p_l, the question whether p_f is less than $\frac{1}{2}$ has no definite answer, implying that personal preference allows either a design or chance explanation to be given.

The planets are an example of what Dembski calls a *probabilistic resource*, partially defined as the number of opportunities for an event to occur by chance [1]. Here the planets determine the number of opportunities for life to arise (and be detected) in the known universe.

The origin of life by chance is assumed to operate according to the following *chance hypothesis*: An organism with a million base pairs, each with four possible nucleotides, arose by chance; the nucleotide for each base pair is stochastically independent and each nucleotide is equally likely; the base pairs randomly assemble only once early in the history of a suitable planet, and if the base pairs assemble in a sequence corresponding to a functional organism, the organism is considered to be functional. This *simpliciter* chance hypothesis is given under the definition of **H** in Appendix B, and was previously referred to as the *simpliciter* understanding. The procedure is to calculate the probability of a functional organism under the assumption of the chance hypothesis just stated. If the probability is less than $\frac{1}{2}$, one infers design, otherwise one retains chance. The procedure is formally developed in Appendix B.

In section 3 our *evolutionary* understanding indicated the probability for an organism to be functional $p_e \geq N_{pl}\, p_{ev}\, 4^{-1000000}$, where p_{ev} is the probability of evolution from a potentially functional progenitor organism to a functional modern organism.

As before, uncertainty in the numbers N_{pl} and p_{ev} leaves a lot of room for personal preference. In addition, as in the fable where the derelict calculator is in a state of decay, one expects partial knowledge about numbers based on the remote past: the million base pairs and p_{ev}.

The probabilistic resources (the planets) remain the same as discussed previously, but the *chance hypothesis* changes to: A progenitor organism with a million base pairs, each with four possible nucleotides, arise by chance. The nucleotides for each base pair are stochastically independent and each nucleotide is equally likely. The base pairs randomly assemble only once early in the history of a suitable planet. If the assembly corresponds to a functional organism, the organism is considered to be functional. The progenitor then has a chance to evolve by gradual or punctuated polyphyletic Darwinian evolution to a modern functional organism with one billion base pairs. This evolutionary chance hypothesis was previously called the evolutionary understanding.

Because of various uncertainties, e.g. partial knowledge of both the present and the past, different assumptions for the probabilistic resources and the chance hypothesis, both design and chance explanations can be offered by an individual or the scientific community.

The *simpliciter* chance hypothesis involves the random assembly of the functional modern organism, while the evolutionary chance hypothesis only posits the random assembly of the progenitor organism. Should the evolutionary chance hypothesis be considered to be superior or inferior to the *simpliciter* chance hypothesis? This depends on the plausibility of the causal story of evolution: the new element in the evolutionary chance hypothesis. Appendix A makes it clear that p_{ev} could in theory attain any value between 0 and 1. It is instructive to analyze the two limiting cases.

In the case of impossibility ($p_{ev} = 0$) evolution is manifestly eliminated as a valid process, and the evolutionary chance hypothesis should not be considered as a basis for the design inference. Numerous computational demonstrations of the viability of evolutionary mechanisms in simplified hypothetical scenarios over the last decade suggest that evolution is possible ($p_{ev} > 0$).

In the case of certainty ($p_{ev} = 1$) evolution is clearly established as a valid process, and only the evolutionary chance hypothesis should be considered as a basis for the design inference. (Note that even in this case, $p_e \geq N_{pl}\, 4^{-1000000}$ so that both the design and chance explanations can be offered. If the progenitor is sufficiently unlikely to arise by chance even certain evolution will not enable a chance explanation.) As in the fable, evolution, though plausible, is not expected to be demonstrated definitively[4] ($p_{ev} = 1$). In fact, one should

[4]If $p_{ev} = 1$ a potentially functional organism will necessarily lead to a functional modern organism. This implies that no catastrophe can bar this process.

not demand necessity of a scientific theory [2]. Unless p_{ev} is very near to unity, i.e. there is a strong reason to prefer the evolutionary above the *simpliciter* chance hypothesis, the evolutionary chance hypothesis is not necessarily superior. This is very different from the situation in physics, where, e.g., the Einsteinian understanding is clearly preferred over the Newtonian one. Why then does there appear to be a bias in the current scientific literature in favor of evolution? This is the subject of the next section.

In this section we sketched two different chance hypotheses, each of which allows either a design or chance explanation to be offered. Uncertainties do not permit any clear preference. The *simpliciter* chance hypothesis cannot be regarded as an *earlier* understanding with the evolutionary chance hypothesis a *later* understanding, although the history of evolutionary biology might give this impression. Both chance hypotheses have to be considered and tension exists between design and chance *both* within the same chance hypothesis, and between different chance hypotheses. In fact there are four quadrants of possible positions: *simpliciter* and design, *simpliciter* and chance, evolutionary and design, and evolutionary and chance[5]. Perhaps most instructive is to view the positions as two pairs: within each chance hypothesis there are two possible explanations. Conjoining "*simpliciter*" and "design", or "evolutionary" and "chance", as though there are only two possible positions, is not admissible. However, in order to understand the current biological debate, it is instructive to note three important scientific theories. We denote by "Intelligent Design" the theory that evolution did not take place, and that the probability of an organism to arise by chance is small, i.e. $p_f < \frac{1}{2}$. Intelligent Design hence is a *simpliciter*–design position. It rejects the chance hypothesis of evolution, and hence does not consider either the evolutionary–chance or evolutionary–design positions. "Darwinian Evolution" denotes the theory that evolution did take place, and that the probability of an organism to arise by chance is large, i.e. $p_e > \frac{1}{2}$. Darwinian Evolution is an evolutionary–chance position. It rejects the *simpliciter* chance hypothesis, and hence does not consider either the *simpliciter*–design or *simpliciter*–chance positions. These two theories in many ways capture the current debate about biological origins, and enables one to understand the history of the biology of the origin of life. Because Intelligent Design rejects the chance hypothesis

[5]Here we deviate from pp. 50–52 and 222–223 of Dembski [4] where chance explanations arising from *all* chance hypotheses should be eliminated before design can be inferred. We consider the single chance hypothesis which is regarded as most accurate, and then infer chance or design based on it. The reasons for this difference between our work and that of Dembski will be detailed elsewhere.

of evolution and Darwinian Evolution accepts evolution, they are logically contradictory theories. There is also an evolutionary–design position, which is the same as Darwinian Evolution, except that the probability of an organism to arise by chance is small, i.e. $p_e < \frac{1}{2}$. For want of a better name, we will refer to this scientific theory as "Theistic Evolution". Darwinian Evolution and Theistic Evolution are logically contradictory scientific theories as the former contains chance, and the latter design.

We restrict our analysis to the (very specific) *simpliciter* and evolutionary chance hypotheses discussed earlier, and accordingly the narrow definitions of the scientific theories mentioned above, for concreteness and simplicity. More realistic chance hypotheses beyond these would consider multiple genera of organisms of various levels of sophistication (not only one organism), each occurring in numerous (near) copies, with DNA base pairs interacting via mRNA as part of a cellular structure involving proteins and enzymes, etc. Attempts at more realistic chance hypotheses and calculation of probabilities have been made [3].

One should seek to formulate a realistic chance hypothesis incorporating currently known scientific knowledge. Even if the chance hypothesis is simplified, it should contain essential features of known scientific knowledge. Even so, some elements of the chance hypothesis may not be scientifically well established. This allows the possibility that more than one chance hypothesis needs to be considered, as in the two chance hypotheses we explicated. In general, there are various chance hypotheses, and in each chance hypothesis a chance or design explanation is possible (Fig. 1). The chance hypothesis should be sufficiently clearly and unambiguously stated to enable the probability relevant to the design inference to be calculated accurately in principle. In the two chance hypotheses we considered, we endeavored to state them clearly. Even though the chance hypothesis should be stated clearly, the scientific knowledge needed for an accurate calculation of the probability relevant to the design inference may still be lacking. In the chance hypotheses we considered, the values of p_l, N_{pl} and p_{ev} were not accurately known. To summarize, the probability should exist, but may not be unique. For the design inference, it is not necessary to know the probability relevant to the design inference (p_f or p_e) accurately. It is sufficient to know whether the probability is less than or larger than $\frac{1}{2}$.

Figure 1: Possible different probabilities relevant to the the design inference that can be calculated for various chance hypotheses. Chance or design is inferred depending on whether the probability is larger or less than $\frac{1}{2}$.

5 The process of biological discovery: *simpliciter* versus evolutionary chance hypotheses

As in the fable, design as a scientific theory was initially (before Darwin) pre–eminent in history (see the writings of William Paley). Science can effect progress beyond the design explanation by reverse engineering, i.e. reconstructing *how* the various components of life were engineered by the designer, which in many ways represents the day to day business of the laboratory scientist. Reverse engineering, as construed here, restricts itself to understanding how the design was assembled and works, i.e. strictly the engineering aspects. It does not include issues related to the designer himself, e.g. why the design was designed, based on which principles the design was made, and what we can know about the designer from his design. These issues will be referred to as the "causal story". Because reverse engineering, and the causal story about the design, are distinct issues, the design explanation and reverse engineering stop short of asking all questions that are of scientific interest. Particularly, reverse engineering pales somewhat in comparison with asking what the causal story of the origin of life is. The design explanation before the time of Darwin was partially based on the improbability of randomly assembling the various components of a living organism, analogous to our experience that when the components of a man–made machine are randomly interchanged, one expects the machine to be inoperative. There was essentially no causal story known. Similar to the operational calculator of Germanicon, the designer has not been found, at least in a scientifically verifiable way, with few clues as to the causal story of the design of life.

The need for progress in science then leads to the formulation of an alternative causal story, e.g. a naturalistic evolutionary understanding à la the fable. The objective here is not to delineate exactly why and how the shift from design to evolutionary understanding happens, as happened at the time of Darwin, but to point out that the hunger for progress in science demands vigorous research on the causal story of the origin of life (whether via evolutionary means, or, presumably, arising from the actions of the designer). This, and the lack of a causal story for design, then led to the development of an evolutionary causal story.

Whichever of the causal stories are making most progress consistent with the experimental data will be in ascendancy in the literature. This explains the sociological fact of the overwhelming presence of evolutionary research in the biological literature since Darwin. We are not claiming that the need for a causal story is the *only* reason for the existence of the literature, but that it is an important reason[6].

Ascendancy in scientific publishing will translate to eminence in the popular media and museums of natural history, where biological knowledge must inevitably be compressed to a few sound bites to satisfy the demand for a causal story. However, one would *not* be warranted to conclude from the predominance of the volume of evolutionary research, and accordingly the size of the community of discourse, that evolution is the only chance hypothesis that needs to be considered.

This follows because, as elucidated in the previous section, and exemplified in the fable by the status of research on calculators between the thirty second and thirty fifth centuries, evolution is far from having been demonstrated definitively. The *simpliciter* chance hypothesis still should be considered as science progresses. The recent literature of the Intelligent Design movement since 1991 [4] strikingly espouses the notion of design, often within the context of the *simpliciter* chance hypothesis.

One accordingly has to issue a warning to the historians of biology: Volume of publications even over a period of centuries does not necessarily imply that only a certain scientific chance hypothesis should be considered. Admittedly, volume of publications, especially over centuries, often *do* imply that only certain theories need to be considered. For example, one would be justified to believe that the dominance of a round earth versus a

[6]Another important reason for ascendancy in the literature is that there is a causal story that is held by a majority of active researchers for paradigmatic, philosophical, demographic or historical reasons.

flat earth in current scientific literature means that round earth theories are the only ones that need to be considered. However, the implication does not obtain for the origin of life.

Although *simpliciter* and evolutionary chance hypotheses should both be considered and should both have been considered for the past 150 years, because both hypotheses contain elements that are not scientifically well established, philosophers of science have rightly looked for criteria by which one hypothesis may be given superiority over the others. An attempt is made to restrict the various chance hypotheses (Fig. 1) to one. Examples are the logical positivist program of verificationism, Karl Popper's falsificationism [2], and Thomas Kuhn's [5] and Larry Laudan's [6] paradigm shifts. It is not our focus in this work to arbiter between these programs, and endorse or reject any of them. Elements of these programs may help one individual to adopt one chance hypothesis, and another individual another hypothesis. In addition, the programs often attempt to also describe how the scientific community as a whole can attempt to adopt only one chance hypothesis. Even though we shall not take a position on whether and how a superior chance hypothesis can be chosen, the discussion in the next section specializes for simplicity to the case where such attempts are possible.

6 Chance scientific theory = chance hypothesis + mode of explanation

As a preliminary, we clarify here our earlier distinction between chance processes and regularity. A regularity maps from event[7] A to B by some explicitly specified rule (e.g. a law). It is a detailed deterministic one to one map. A chance process is different from this. It lacks a detailed deterministic description. Instead, it posits a chance hypothesis under which the event occurs. Since it is a chance hypothesis, the mapping from event A to B is not deterministic, which implies that it is not a mapping from one to one, but from one to many. That we describe the event B by a chance process does not have to entail that the process has no possible deterministic description. It only entails that we simply do not know enough about the process to constrain it to a one to one mapping.

In section 4 three scientific theories were mentioned: Intelligent Design, Darwinian Evolution and Theistic Evolution. These are theories of chance processes which incorpo-

[7]In the case of quantum theories, the word "event" should be replaced by "wave function".

rate both a specific chance hypothesis and an mode of explanation (design or chance). The reason for the presence of a mode of explanation is that the event described by a chance hypothesis can be explained as either due to chance or design. Theories of chance processes are different from the theories of regularities (e.g. the laws of nature that are found in Einsteinian relativistic and Newtonian physics[8]). The skeleton of a theory of chance processes involves *both* a chance hypothesis and a mode of explanation, while that of a theory of regularities is just the detailed deterministic map (e.g. the laws of nature).

For theories of chance processes one attempts to choose the superior chance hypothesis, as discussed in the previous section, and *given* that chance hypothesis, one must calculate the probability relevant to the design argument, and based on that decide whether design or chance is inferred. Because the probability is taken not to be known to be above or below $\frac{1}{2}$ in view of current scientific knowledge, as elucidated for the explicit scientific theories in section 4, both design and chance can be inferred. There are hence two scientific theories that are concurrently held by the scientific community. Because the one theory corresponds to design, and the other to chance, within the *same* chance hypothesis, the theories are logically contradictory. It is hence possible for two logically incompatible scientific theories to be adopted concurrently by the scientific community.

For theories of regularities one attempts to choose the superior laws of nature somehow, along the same lines as the way the superior chance hypothesis is chosen for theories of chance processes. Once the superior laws of nature are chosen only one scientific theory is adopted by the scientific community, so that two logically incompatible scientific theories are not adopted. For example, Kuhn's account of the progress of science, which is grounded in a historical study of the Copernican, Galilean, Newtonian and Einsteinian revolutions in physics, would say that there is only one scientific theory that can be adopted by the scientific community during non–revolutionary periods (his principle of *incommensurability*) [5].

Theories of chance processes behave in a very different way to theories of regularities, because their skeleton is different: the former has in addition a mode of explanation, which is absent in the latter. The recognition of the existence of this mode of explanation as part of the methods of science may be viewed as the pivotal insight of Dembski [1, 4]. As elucidated in the next section, the mode of explanation is *rational* and hence should be

[8]Quantum theories contain elements of both chance and regularity, which go beyond the current work. The evolution of the wave function is described by deterministic equations, but the positions and momenta of particles are probabilistically distributed according to a distribution determined from the wave function.

incorporated in the methods of science.

We have argued that for theories of chance processes two logically contradictory scientific theories can coexist. The fundamental reason for the coexistence is that it is impossible to calculate whether the probability relevant to the design argument is above or below $\frac{1}{2}$ within a single chance hypothesis. However, one can certainly forsee a time in the future when the probability may be calculated accurately enough to allow only one of the explanations to be offered. In this sense the period of coexistence of logically incompatible scientific theories should be viewed as temporary: a lengthy period of metastability. Off course, if the probability relevant to the design inference can at the present time be accurately calculated to be above or below $\frac{1}{2}$, there need be no period of coexistence of logically incompatible scientific theories.

As an afterthought it is interesting to mention some of the sociological consequences of the coexistence of two logically contradictory scientific theories. Both theories develop an entrenched orthodoxy, because the theories are logically inconsistent, and the stage is then set for dirty tactics by constituencies and lobbies. This explains the current impasse and hostility between design and chance on the academic level, and spills over into the popular debate, including the political question of which theory of origins should be taught at educational institutions. In another context, similar dynamics were observed by Kuhn when he concluded that the orthodoxy of the one paradigm attempted to thwart the visionaries of another paradigm before, during and the after the period of revolution [5]. It should be emphasized that the entrenched orthodoxy on both sides is a corollary of the sustained coexistence of logically incompatible scientific theories. We hence do not explain the coexistence of two orthodoxies merely by the existence of strong religious (theistic) constituencies, in the case of design, and atheist (naturalistic) constituencies, in the case of chance, as is often done. This not to deny the power of these constituencies, but simply to say that they cannot sustain those orthodoxies over the long term in a society that values scientific progress. This is related to Kuhn's claim that paradigms cannot merely be upheld by social pressure, except in the short turbulent times of revolution, if scientific progress is a value of society [5].

7 Design and chance: separate but equal

Since our primary interest in the origin of life is not in regularity as a mode of explanation, we consider only chance and design. The logic of the design inference is such that the hypothesis of chance is assumed unless the law of small probabilities shows that the event did not occur by chance, i.e. it was designed. As a mode of explanation, chance is hence prior to design and has priority over design. This ruling for priority of the chance explanation is not logically necessary, but is due to quintessential methods of science. It is demanded by Occam's razor and the principle of testability. Those who presuppose design, should hence be regarded as unscientific, but not incoherent. To see this let's put design prior to chance. Any event that occurs is *prima facie* ascribed to design. This means that each event possesses some quality (entity) beyond the event that is attached to the event. This is at the very least the (non–vacuous) quality of being designed, but usually involves additional information about possible designers. Qualities range from the mundane (the event was planned by my aunt) to the esoteric (the event is due to cosmic consciousness or due to a member of the pantheon of Athenian gods). Occam's razor means that we should not multiply entities beyond necessity, i.e. we should not introduce these non–essential qualities. The principle of testability demands that it must be possible to test scientific theories. The more esoteric qualities cannot be tested experimentally. (Although it should be clear that the presupposition of design can lead to the assumption of egregious non–essential qualities, even ascribing an event to chance leads to the assumption of certain qualities. The event is assigned the quality of "having arisen by chance".) Another reason for putting chance prior to design is that a scientist would rather presuppose that an event was a coincidence and not that it was designed. For example, meeting a familiar person in a supermarket should first be considered to be a coincidence. If one immediately surmises that the meeting was planned, you could be guilty of paranoia [1]!

Putting chance prior to design is hence demanded by the methods of science. This is not equivalent to saying that chance is superior to design as a mode of explanation.

A commonly found way to give superiority to the chance explanation is to accept design as a mode of explanation, but then to add an additional clause requiring a causal story in the case of a successful design inference. For example, one may add the clause (question): Was the designer part of nature? A computer, robot, human or animal designer would fall in this category. In case the question is answered affirmatively, design (and also

a designer that is part of nature) is inferred. In case the question is answered negatively, one goes back to consider chance, does further investigation leading to a different chance hypothesis, does not infer chance or design, go back to supposing chance, and progress in a loop until one (hopefully) eventually fixates on either design (and a designer that is part of nature) or chance. In this way one explores many different chance hypotheses.

The additional clause is of course *ad hoc*. It is added in order not to allow the deduction that the designer is not part of nature (e.g. God), which amounts to nothing more than an *ad hoc* requirement stemming from metaphysical naturalism (that there exists nothing beyond nature) [4]. Such metaphysical presuppositions cannot be incorporated into the methods of science[9]. A simple example also demonstrates why this additional clause leads to irrational results. Consider the book "The Cloud of Unknowing" written by an unknown medieval author. The answer to the question "Was the designer part of nature?" is strictly speaking negative, since the author is unknown. One now reconsiders chance, does further investigation, but still do not identify the author, does not infer chance or design, and proceed in a (potentially) infinite loop. This situation is clearly unsatisfactory in that the existence of the book should have led one to infer that it was designed.

Superiority of chance in biology can stymie research, e.g. presuming that so–called "junk DNA" is the left–overs from evolutionary mutation and chance assembly that has no function, and strongly resisting the possibility that junk DNA has a function, would stymie the research that indicated that junk DNA is not junk.

After delineating an (incorrect) way how the chance explanation can be given superiority over design, and giving an example of how superiority of chance could stymie research, we now outline why these explanations should be on an equal footing. It follows from the *rational* principle that *if a conjecture $\mathcal{C}$ has probability less than half of describing an unknown event happening by chance, then the logical complement of $\mathcal{C}$ should be conjectured* [1]. For example, consider the unknown result of a fair coin tossed two times. If the conjecture $\mathcal{C}$ is that two heads will appear, which has a probability of $\frac{1}{2^2} = \frac{1}{4}$, then the logical complement of $\mathcal{C}$, i.e. that either one head and one tail or two heads will appear, should be conjectured. This is *rational* since one is more likely to be correct if the logical complement of $\mathcal{C}$ is conjectured. Dembski shows pedantically how similar principles imply

[9]Science cannot be done without certain presuppositions (ansätze). Amongst these are the *methods of science*, referred to before. However, these presuppositions must be minimized.

that chance and design are on an equal footing [1]. A compact way to motivate this is to replace $\mathcal{C}$ with "chance" in the above principle: "If a chance conjecture has probability less than half of describing an unknown event, then design should be conjectured".

Having outlined that design is a *rational* explanation, it is evident that it is a mode of explanation that should be included as a method of science. Because design is the logical complement of chance, inferring design means that the explanation that the event is due to chance (under a specific chance hypothesis) is negated. One of the reasons for the failure of chance may be that the specific chance hypothesis, in which tremendous scientific research was expended, is unreliable. Lack of success of the chance hypothesis is immediately penalized by a design explanation. The tension is that the design inference may be viewed as a killer of scientific research in that it may result in the loss of the chance hypothesis in which so much energy was wasted. Moreover, the design inference has not replaced the lost scientific research with new scientific research, and may hence appear counterproductive. However, this is not a sufficient reason to exclude the design inference from the scientific domain, since the design is a perfectly rational explanation to offer. After all, discovery of Kant's book *Kritik der Reinen Vernunft* should not distract us from concluding that it was written by an author (designed) even though we may have expended tremendous scientific research trying to explain how the book could have arisen by purely naturalistic means (chance).

Our usual experience in science is that it is not sufficient just to criticize a scientific theory. The scientific theory needs to be replaced by a superior theory which is accompanied by new scientific research. Unless a superior theory is formulated, the imperfect theory remains the consensus of the day. In our case, the initial starting point is the chance hypothesis, the endpoint is the design explanation, but in this case it may not be accompanied by new scientific research. When chance is replaced by design, it is hence very different from the usual case when a superior theory replaces an imperfect theory. If we go back to section 6 we see that the reason for the different behavior is that the progress from chance to design just described is a change of *mode of explanation* in a theory of chance processes, while the progress between the imperfect and superior theory described here is a change of either *chance hypothesis* in the case of theories of chance processes, or laws of nature in the case of theories of regularities.

Having outlined why design and chance should be treated *equally*[10], i.e. the magical number "$\frac{1}{2}$", there is still the nagging doubt that the scientific method may prefer chance. The worry is that if one makes the design explanation too easy, one commits oneself to entities beyond necessity, which should be suppressed by virtue of Occam's razor. But this is in fact a fallacy. Since design is simply defined as the logical complement of chance, there are no additional entities introduced by a design inference.

In this section we argued that chance has *priority* but not *superiority* over design as a mode of explanation. A very common misconception is that because chance is *prior* to design, i.e. one first formulates a chance hypothesis to explain an event, it follows that chance is the explanation for the event, i.e. chance is highly *superior* to design as a mode of explanation. This is tantamount to saying that because a chance hypothesis has been posited to explain an event, the event can be explained by chance. This misconception arises from equating *priority* and *superiority*, which we have argued to be incorrect.

8 Epitome

In section 2 we discussed the concrete example of the origin of an organism with a billion DNA base pairs and outlined how the design argument proceeds. The base pairs are taken to assemble randomly into a functional organism, referred to as the *simpliciter* chance hypothesis. In section 3 we introduced the causal story that the organism evolved from a progenitor organism with a million base pairs, called the evolutionary chance hypothesis. In section 4 we indicated that there are four positions, corresponding to either the design or chance explanation, in either the *simpliciter* or evolutionary chance hypotheses. These positions are adopted based on probabilities which are uncertain. The *simpliciter* and evolutionary chance hypotheses both have to be considered, as there is no clear preference for either. In section 5 we argued that the *apparent* predominance of the evolutionary chance hypothesis is partially explained by the need for a causal story. Section 6 shows that the skeleton of theories of chance processes (biology of the origin of life) is very different from

[10]Treating design and chance equally may seem to be at odds with statistical hypothesis testing where 95% or 99%, not 50% confidence levels for rejecting a chance hypothesis are commonly set. High confidence levels are set in order to control false positives in the scientific literature, because one wants to reduce the risk of publication of incorrect results. However, it is perfectly rational to make a judgment based on a 50% confidence level.

that of a theories of regularities (e.g. Newtonian physics). In the case of chance processes, two logically contradictory scientific theories coexist during an extended period of metastability. An entrenched orthodoxy is implied by the coexistence of these theories. Section 7 explores and refutes three attempts to give superiority to chance: (1) requiring a causal story in the case of a successful design inference, (2) viewing design as counterproductive because it does not replace scientific research with new research, and (3) claiming that a design explanation commits one to entities beyond necessity.

A careful reading, leading to corrections and new additions to this work, by Michael Francisco, Dr. William Powers, and members of the Megaviews Forum, is acknowledged.

References

[1] W. Dembski, "The Design Inference: Eliminating Chance by Small Probabilities", Cambridge University Press, 1998, ISBN 0-521-62387-1.

[2] K.R. Popper, "The Logic of Scientific Discovery", Routledge Kegan & Paul, ISBN 0-415-07892-X; "Conjectures and Refutations: The Growth of Scientific Knowledge", Routledge Kegan & Paul, 1962, ISBN 0-710-06507-8.

[3] H.P. Yockey, "Information Theory and Molecular Biology", Cambridge University Press, 1992, ISBN 0-521-35005-0.

[4] W. Dembski, "Intelligent Design", Intervarsity Press, 1999, ISBN 0-8308-1581-3.

[5] T.S. Kuhn, "The Structure of Scientific Revolutions", University of Chicago Press, 1962, ISBN 0-226-45807-5.

[6] L. Laudan, "Science and Values : The Aims of Science and their Role in Scientific Debate", University of California Press, 1984, ISBN 0-520-05267-6.

A Appendix: Evolution as cumulative probability

Consider the following simple mathematical algorithm:

(1) Start with the null (empty) strings S_i where $i = 1, \ldots, n$.

(2) Randomly choose the string $P_j =$ "0" or "1", for each $j = 1, \ldots, n$.

(3) Concatenate the string S_i to the string P_i to form the string S_iP_i for each $i = 1, \ldots, n$.

(4) Are any of the strings S_iP_i composed only of 1's? If so, continue. Otherwise set all S_iP_i equal to null strings and halt.

(5) Take one of the strings with all 1's. If the string is of length m halt. Otherwise assign S_i to be this string for each $i = 1, \ldots, n$. Go to (2).

In this algorithm, we start with n strings, each of which is randomly assigned a value of "0" or "1". If any of the strings is "1", we choose this string and make $n-1$ copies of it. To each of these strings consisting of "1", we now add a randomly assigned value of "0" or "1". If any of these strings is "11", we choose this string and make $n-1$ copies of it. The process continues until a string of length m consisting only of 1's is made.

The probability of constructing a string consisting of m 1's by trying out n strings can be calculated to be

$$p_{ev} = (1 - \frac{1}{2^n})^m.$$

If one tries out as many strings as possible, one is certain to construct a string of length m consisting of only 1's ($p_{ev} \to 1$ as $n \to \infty$). It is also very difficult to construct a very long string of 1's ($p_{ev} \to 0$ as $m \to \infty$). It is critical that the eventual string of 1's is a meaningful sequence, not just a random sequence of 0's and 1's. A string "1111111111" is a meaningful sequence since it can be specified by side information like "repeat '1' 10 times". As discussed by Dembski, such a string of 1's is *specified* (meaningful) [1].

The algorithm exemplifies biological evolution. A string represents a sequence of DNA base pairs. As the strings become larger, more complicated organisms are built. If an organism that can survive under natural selection (a string of only 1's) is found, that organism makes n copies of itself and dies. The copies are its offspring. A new base pair is now added to each offspring, with a value that is determined randomly (mutation). If no organism that can survive under natural selection is found, the organisms are destroyed. The process terminates when an organism with m base pairs that can survive natural selection is found. The probability of making such a "functional" organism is p_{ev}. In the same sense that a sequence of 1's is a meaningful string, the organism that survives natural selection has special meaning: it is functional. Note that the two main tenets of evolution, mutation and natural selection, are present in the algorithm.

The probability of producing a string of m 1's by random coin tosses is $1/2^m$, a much

smaller number than the probability p_{ev} of producing a string of m 1's by the algorithm (using cumulative probability)[11]. Cumulative probability is the miracle of evolution.

In the main text we are interested in evolution that starts with a potentially functional progenitor organism. To accommodate this the algorithm needs to be changed to start with a string of 1's which copies itself to a n strings, instead of starting with n null strings.

B Appendix: The *simpliciter* design argument for the origin of life

In section 2 we sketched the *simpliciter* design argument needed to establish the design of life. Let's briefly recapitulate. Consider a fairly complex organism with a billion DNA base pairs, each of which can take four possible values. The relevant probability to infer design from is the probability of an organism to be functional given an *ad hoc* sequence of base pairs p_f. The probability of finding functional organisms $p_f = N_{pl}\, p_l$, where p_l is the probability of an organism to be a functional organism *on earth*, given an *ad hoc* sequence of base pairs, and N_{pl} is the number of suitable planets on which intelligent life arose.

In this appendix we show that p_f is the relevant probability to infer design from and that it has the value $N_{pl}\, p_l$ we claimed. This is done by formulating the problem in the language of Dembski [1]. The reader is referred to this work for an exhaustive explanation of terms.

As a preliminary, we note the distinction between an existent functional organism in nature, and the abstract knowledge of functions. The former is the realm of events, and the latter is the realm of patterns. We now summarize the generic design argument and then make it explicit for the origin of life.

Suppose a subject S has identified a chance hypothesis **H** that could be responsible for an event E. Suppose further that S has identified (1) a probability measure P that estimates likelihoods with respect to the chance hypothesis **H**, (2) side information **I**, (3) a bounded complexity measure φ that characterizes S's problem solving ability, and (4) probabilistic resources Ω. Now S identifies a pattern D that describes the event E, and calculates the probability p of the event corresponding to the pattern, D^*. Then S identifies

[11] $1/2^m \leq p_{ev}$ for all n.

the event corresponding to the pattern incorporating the probabilistic resources, D^*_Ω, and calculates the associated probability p_Ω.

S	Biologist
H	There are one billion base pairs each consisting of four possible nucleotides. The base pairs are chosen stochastically independent of each other and each nucleotide is equally likely. The base pairs randomly assemble only once early in the history of a suitable planet. If the base pairs assemble in a sequence corresponding to a functional organism, the organism is considered to be functional.
P	The frequentist probability measure
E	An existent functional organism
D	A specific known set of functions
D^*	An organism with this specific known set of functions
p	The probability of finding an organism with a specific set of known functions
I	Knowledge of different possible functions
φ	Complexity of choosing a specific set of functions from the knowledge of different possible functions
Ω	Information about how many sets of different known functions there are (Ω_f) and how many suitable planets with intelligent life there are (Ω_{pl})
N_f	Number of sets of different known functions
$N_{l\|f}$	Number of functioning organisms there are for each different set of known functions
N_l	Number of functioning organisms
N_{pl}	Number of suitable planets with intelligent life
D^*_Ω	Any organism that can perform known functions arising from any suitable planet with intelligent life
p_Ω	The probability of finding a functioning organism

Since our interest is in finding life and not a specific life form, D^*_Ω interests us, not D^*. Accordingly, the relevant probability to infer design from is $p_\Omega \equiv P(D^*_\Omega|\mathbf{H})$, and not $p \equiv P(D^*|\mathbf{H})$. From these definitions of p_Ω and p,

$$p_\Omega = N_{pl} \times N_f \times p$$

$$
\begin{aligned}
&= N_{pl} \times N_f \times N_{l|f} \times 4^{-1000000000} && (\text{using } p = \frac{N_{l|f}}{4^{1000000000}}) \\
&= N_{pl} \times N_l \times 4^{-1000000000} && (\text{using } N_l = N_f \times N_{l|f}) \\
&= N_{pl} \times p_l && (\text{using } p_l = \frac{N_l}{4^{1000000000}}).
\end{aligned}
$$

Now p_Ω is identical to p_f, so that $p_f = N_{pl} p_l$ as claimed.